Benchmark Phonics Intervention: Midwest Study Results

Abstract

Even with a comprehensive English Language Arts core program that is well taught, there will be students who need more instruction and more practice while learning to read. This is now particularly true because of the COVID-19 pandemic. *Benchmark Phonics Intervention* (BPI) was developed to provide Tier 3 instruction for students two or more years behind in reading ability. In a Midwest district during the 2021–2022 school year, a quasi-experimental study was conducted comparing instruction with BPI to Tier 3 intervention instruction traditionally provided by the district. In the year-long study, 66 students in Grades 1–5 who received BPI instruction across the school year were compared to 125 students in demographically similar elementary schools. Results showed that based on i-Ready scores, the two groups of students were statistically similar at the beginning of the year and statistically significantly different at the end of the school year, in favor with of the BPI students. This indicates students who received the BPI instruction across the school year made more progress than the students in the comparison group. Three program assessments used by the treatment students supported the findings by showing statistically significantly different growth and very large effect sizes.

Introduction

Benchmark Phonics Intervention (BPI) is a Tier 3 phonics intervention program, built by Benchmark Education Company (BEC), that provides systematic, explicit phonics instruction to students who are two or more years below level, students with dyslexia, and students with other learning disabilities who struggle with learning to read. During the 2021–2022 school year, BPI was implemented as part of this study in a school district in the Midwest. Three elementary schools were identified as the treatment group, and four schools in the same district served as the comparison group. This district already had a very strong intervention program, with interventionists that were well trained and with a focus on prevention in the early grades. This report describes the study, including the participants, methods and guiding research question, assessment descriptions, and the results of this study.

Participants

With any new program, there was interest in finding districts willing to support a study to examine the possible program results. The district that agreed to participate in the study showed interest in trying a new phonics intervention program and was interested in this particular phonics program because it was authored by Wiley Blevins. People in the district were already familiar with Blevins's work regarding the teaching of phonics. The district's contact person was part of the Title I staff for the district and was very familiar with the current interventions used by the district.

The contact person was able to identify three elementary schools that were willing to participate as the treatment group. This initially meant that for the 2021–2022 school year, current intervention instruction would stop and BPI would be used for students in need of Tier 3 intensive phonics instruction. The contact person also identified four additional schools, demographically similar to the treatment schools, to act as the comparison group. Training and follow-up meetings for the BPI study were provided via Zoom meetings, influenced by the ongoing pandemic.

As mentioned before, this district was focused on early intervention to try to prevent students from falling behind, especially in the early grades where students had missed critical instruction in early

reading fundamentals because of the pandemic. As a result, students participating in the study included Grade 1 students. Grade 1 students are not usually part of the target audience for BPI, but with the talented intervention teachers, they were able to make it work. This would still not be recommended by BEC for the BPI program because the instruction is designed for students two years or more below grade level.

When conducting a study in a natural setting, such as in elementary schools, the ultimate responsibility for the students' academic progress and well-being rests with the school and its staff. At the midyear point of this study, the interventionists wanted to be able to move students into or out of (to other intervention or to discontinue Tier 3 instruction) or modify the BPI instruction, based on their expertise and relationships with the students. Additionally, students who moved into the district and/or into a study school needed to be assigned to Tier 3 intervention classes, that happened to be using BPI instruction, for the second half of the school year. Since the study was designed to examine a program for students with long-term intervention needs, the students who received BPI instruction for only half of the school year (n = 50) were excluded from the final treatment group, called BPI Full Year.

The final number of students who participated in the study included 125 students in the Comparison group and 66 students in the BPI Full Year group. The demographics of the BPI Full Year group, based on the schools in the study, include 78% free/reduced price lunch, 64% White, 13% Hispanic, 13% Black, 10% Two or more races, and a very small number of students who identify as American Indian, Asian, and Pacific Islander.

Methods and Guiding Research Question

This study used a quasi-experimental design with district-determined intact student groups. Students in Grades 1–5 who participated in the study were determined to be eligible for intervention instruction based on the school and district guidelines. Examination of the beginning of year (BOY) scores demonstrated statistical equivalence between the BPI Full Year and Comparison groups at the start of the study.

The research question that informed this study was based on progress that could be made using the *Benchmark Phonics Intervention* program for the Tier 3 pullout instruction compared to a group of students who received the district's usual intervention instruction, also using pullout instruction. The Comparison group received intervention using the currently available materials, and each of the four schools used similar intervention instructional materials. The classroom core ELA instruction for students in both groups was similar.

Assessments

There were four assessments used during this study. These assessments are briefly described below.

- i-Ready Diagnostic is an adaptive assessment that adjusts its questions to suit students' needs. This assessment was one of the major assessments that was given to Grades K–5 students in the district. This was the main assessment used to compare the two groups of students in this study.
- Quick Phonics Assessment (QPA) is an informal, criterion-referenced, individually administered diagnostic assessment tool created by Dr. Jan Hasbrouck. QPA is designed to help teachers and specialists quickly and accurately diagnose students' strengths and instructional needs in

phonics and decoding skills and has been successfully used for over 10 years. Alternative forms are available for monitoring student progress at three times across the school year.

- Quick Spelling Assessment (QSA), also created by Dr. Jan Hasbrouck, is designed to be used along with QPA. The primary purpose of QSA is to help teachers and interventionists decide which students should be given QPA and on which Skill Set to start the assessment. Results from QSA can also supplement and expand the information that is obtained from the administration of QPA by providing information about students' encoding and spelling skills.
- The Phonological and Phonemic Awareness Assessment, created by Wiley Blevins, examines a student's range of phonological and phonemic awareness skills in a one-on-one setting. This assessment is designed to determine whether students need or would benefit from more intensive instruction and practice in phonological and phonemic awareness.

The original intent of this study was for all participating students, regardless of group, to take all four of the assessments at BOY and end of year (EOY). As things happened, the Comparison groups decided not to administer the QPA, QSA, or Phonological and Phonemic Awareness Assessment. That left the i-Ready as the assessment used to compare progress across the school year between groups. Within the treatment group, only one school out of the three administered the Phonological and Phonemic Awareness Assessment. Results were still calculated for as many students as possible who took assessments and had scores at each assessment period, BOY and EOY.

Results: i-Ready Assessment Results

The initial thought for analysis of the i-Ready assessment results was to use the independent-samples t-test. When determining if the independent-samples t-test could be used with this data set, outliers were discovered in the data set, as assessed by inspection of a boxplot. There were outliers in both groups (at least one was extreme). Outliers are extreme values in a data set that can distort results.

Based on these findings, the best course of action was to run the nonparametric Mann-Whitney U test. The Mann-Whitney U test is a rank-based nonparametric test that can be used to determine if there are differences between groups using medians, which are less likely to be affected by outliers. The Mann-Whitney U test is often used as the nonparametric alternative to the independent-samples t-test when the data fail the required assumptions.

A Mann-Whitney U test was run to determine if there were differences in the i-Ready scores at BOY and EOY between the students in the Comparison group and the students in the BPI Full Year group. Distributions of the i-Ready scores at BOY and EOY for the two groups of students where similar, as assessed by visual inspection. The data met all four assumptions of the Mann-Whitney U test. The result statements are as follows.

- Median i-Ready scores at BOY for the Comparison group (mean rank = 91.21) and BPI Full Year group (mean rank = 105.08) were not statistically significantly different between the two groups: U = 4,724.00, z = 1.649, p = .099.
- Median i-Ready scores at EOY for the BPI Full Year group (mean rank = 107.85) were statistically significantly higher than for the Comparison group (mean rank = 89.74): U = 4,907.00, z = 2.153 p = .031.

The implications of these results statements are twofold. First, at BOY, the Comparison and BPI Full Year groups were similar from a statistical perspective based on the i-Ready scores. Second, at EOY, these two groups were statistically significantly different based on the i-Ready scores. Knowing the instruction in the school day was similar for all students in each group, the different intervention instruction that students received made a difference in the growth students made on the i-Ready assessment from BOY to EOY. The students in the BPI Full Year group achieved more growth on the i-Ready assessment across the school year than did the students in the Comparison group.

Results: Quick Phonics Assessment Results

The QPA, which measures a student's ability to recognize, decode, and pronounce phonics elements, was administered to 55 students who had scores at both BOY and EOY in the BPI Full Year group. Table 1 shows the specifics for the QPA for the whole group and for students by grade. There were no students who had matched scores at BOY and EOY who were in Grade 3. Understanding the scores for this assessment requires a brief description of the assessment.

The QPA has a series of Skill Sets that starts from the basics, Skill Set 1: Letter Names, and moves through the phonics skills, in order of difficulty, ending up at the last set, Skill Set 13, where students decode words with four syllables. Each Skill Set has points that can be earned for decoding words correctly. At each set there is a threshold, usually 50 percent or more items correct, that a student must reach to continue to the next level. The Skill Set where the student does not reach the threshold indicates where instruction should probably begin. Students do not need to start the QPA at Skill Set 1 unless there is little information about the student. If the student does not reach the threshold on the starting Skill Set, the previous Skill Set should be administered next. Students get credit for the Skill Sets below the Skill Set where the initial threshold is reached.

On the QPA, 242 total points can be earned if each item in the 13 Skill Sets is completed successfully. To understand the growth made by students across the year, the average points earned (total points scored + any points prior to the starting set) at BOY were compared to the average points earned at EOY.

Looking at Table 1, for All Grades (which included 55 students), on average, 70.33 points were earned at BOY. At EOY, the same 55 students earned an average of 135.36 points. The final four columns of Table 1 show the average starting and ending Skill Sets for BOY and for EOY, based on the Skill Set number each student reached.

Table 1. Quick Phonics Assessment Results for BPI Full Year Students

Grade	# Students	Points Earned at BOY	Points Earned at EOY	BOY Average First Set	BOY Average Last Set	EOY Average First Set	EOY Average Last Set
All Grades	55	70.33	135.36	1	4	3	8
Grade 1	24	40.21	109.38	1	2	1	6
Grade 2	25	82.76	140.32	1	4	3	8
Grade 4	4	124.25	213.25	2	7	8	12
Grade 5	2	168.50	229.50	6	8	8	13

Skill Sets: 1: Letter Names; 2: Letter Sounds; 3: VC and CVC; 4: Common Beginning and Ending Consonant Digraphs; 5: CVCC and CCVC; 6: Silent e; 7: r-Controlled Vowels; 8: Advanced Consonant Sounds, Silent Consonants, and Consonant Digraphs/Trigraphs; 9: Vowel Digraphs, Diphthongs, and Advanced Vowel Sounds; 10: Common Prefixes and Common Suffixes; 11: Two Syllables; 12: Three Syllables; 13: Four Syllables

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the total QPA points earned at BOY compared to those earned at EOY. Three outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of their values did not reveal them to be extreme and they were kept in the analysis. The paired-samples t-test run on the QPA data showed a statistically significant increase in the EOY Total Points scores compared to the BOY Total Points scores: M = 65.036, 95% CI [54.563, 75.510], $t(54)$ = 12.450, $p < .001$. The effect size for the Total Points scores was d = 1.68[1], a large effect size.

Results: Quick Spelling Assessment Results

The QSA provides information about students' encoding and spelling skills and uses the same Skill Sets as the QPA. On the QSA, 272 Total Points can be earned if all items in the 13 Skill Sets are spelled correctly. Table 2 shows the results for the 57 students who had scores at both BOY and EOY. The calculations of Points Earned (Columns 3 and 4) are very similar to the calculations for QPA. The Points Earned is the total number of points earned by the student plus credit for Skill Sets below where the student started and reached the threshold.

Looking at Table 2, for All Grades (which included 57 students), 73.39 points were earned at BOY. At EOY, the same 57 students earned an average of 111.37 points. The final four columns of Table 2 show the average starting and ending Skill Sets for BOY and for EOY, based on the Skill Set number students reached.

Table 2. Quick Spelling Assessment Results for BPI Full Year Students

Grade	# Students	Points Earned at BOY	Points Earned at EOY	BOY Average First Set	BOY Average Last Set	EOY Average First Set	EOY Average Last Set
All Grades	57	73.39	111.37	2	4	3	6
Grade 1	24	48.29	98.08	1	3	2	6
Grade 2	25	78.64	109.40	1	5	3	7
Grade 4	4	111.25	134.75	3	7	4	8
Grade 5	4	153.25	180.00	7	8	7	10

Skill Sets: 1: Letter Names; 2: Letter Sounds; 3: VC and CVC; 4: Common Beginning and Ending Consonant Digraphs; 5: CVCC and CCVC; 6: Silent e; 7: r-Controlled Vowels; 8: Advanced Consonant Sounds, Silent Consonants, and Consonant Digraphs/Trigraphs; 9: Vowel Digraphs, Diphthongs, and Advanced Vowel Sounds; 10: Common Prefixes and Common Suffixes; 11: Two Syllables; 12: Three Syllables; 13: Four Syllables

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the total QSA points earned at BOY compared to those earned at EOY. There were no outliers in the data, as assessed by inspection of a boxplot. The paired-samples t-test run on the QSA data showed a statistically significant increase in the EOY Total Points scores compared to the BOY Total Points scores: M = 37.982, 95% CI [31.174, 44.791], $t(56)$ = 11.175, $p < .001$. The effect size for the Total Points scores was d = 1.42, a large effect size.

[1] Effect sizes were calculated for the paired-samples t-test by dividing the mean difference by the standard deviation of the differences.

Results: Phonological and Phonemic Awareness Assessment Results

The Phonological and Phonemic Awareness Assessment is usually used only when the teacher or interventionist feels there might be an issue with the students' skills. In this case, only students in a small subset of students in the BPI Full Year group completed the Phonological and Phonemic Awareness Assessment. The Skill Sets on this assessment are quite different from the QPA and QSA, but the scoring that was used to look at growth is similar. There are 17 Skill Sets and 170 total points that can be earned on this assessment.

Looking at Table 3, for All Grades (which included 12 students), 64.25 points were earned at BOY. At EOY, the same 12 students earned an average of 135.25 points. The final four columns of Table 3 show the average starting and ending Skill Sets for BOY and for EOY, based on the Skill Set number students reached. It is not clear if the phonological and phonemic awareness instruction during the study period was done as part of the BPI lessons or if additional, more intensive instructional materials provided with BPI were used.

Table 3. Phonological and Phonemic Awareness Assessment Results for BPI Full Year Students

Grade	# Students	Points Earned at BOY	Points Earned at EOY	BOY Average First Set	BOY Average Last Set	EOY Average First Set	EOY Average Last Set
All Grades	12	64.25	135.25	1	12	1	15
Grade 1	10	58.60	130.00	1	11	1	15
Grade 2	2	92.50	161.50	1	15	1	17

Skill Sets: 1: Word Awareness; 2: Identify Rhyme; 3: Produce Rhyme; 4: Clap Syllables; 5: Blend Syllables; 6: Delete Syllables; 7: Segment Initial Sounds; 8: Segment Final Sounds; 9: Segment Medial Sounds; 10: Segment Sounds; 11: Blend Sounds; 12: Substitute Initial Sounds; 13: Substitute Final Sounds; 14: Substitute Vowel Sounds; 15: Add Initial Sounds; 16: Add Final Sounds; 17: Delete Sounds

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the total Phonological and Phonemic Awareness Assessment points earned at BOY compared to those earned at EOY. There were no outliers in the data, as assessed by inspection of a boxplot. The paired-samples t-test run on the Phonological and Phonemic Awareness Assessment data showed a statistically significant increase in the EOY Total Points scores compared to the BOY Total Points scores: $M = 71.000$, 95% CI [54.487, 87.513], $t(11) = 9.463$, $p < .001$. The effect size for the Total Points scores was $d = 2.73$, a large effect size.

Discussion and Summary

This study placed a new phonics intervention program, *Benchmark Phonics Intervention*, in a school district in the Midwest that already had a strong intervention program. The interventionists were highly trained but were being asked to use a totally new program that had to be learned while teaching. This is always a difficult situation. The other piece that made this study just a bit harder was the continuing pandemic that caused the district to close at least once for multiple days because of a rise in infections.

Despite the difficulties, results show that there was a statistically significant difference on the i-Ready assessment between the BPI Full Year and Comparison groups, in favor of the BPI Full Year group. The i-Ready assessment does not assess only phonics, decoding, and encoding, but rather the full ELA spectrum of foundational skills as well as comprehension, vocabulary, background knowledge, etc. The

results indicate that transfer of the skills and knowledge has occurred, enabling the BPI Full Year students to show statistically significantly different growth, compared to the Comparison group using other intervention materials, on the norm-referenced test.

For the criterion-referenced assessments, QPA and QSA, results show the increase of phonics, decoding, and encoding abilities of these students. All students in the BPI Full Year group at the beginning of the year, on average, were working at about the Common Beginning and Ending Consonant Digraphs skill level. By the end of the year, these students were working at the r-Controlled Vowels skill level. The growth between BOY and EOY is statistically significantly different, and the effect sizes are very large: d = 1.68 for QPA and d = 1.42 for QSA. Both effect sizes show about 1.5 standard deviations of movement from the beginning to the end of the school year.

This is the first study of the *Benchmark Phonics Intervention* program. This study will be only the start of an accumulation of results for this program in different locations, with different groups of students in need of Tier 3 intervention for reading. The modifications requested by the interventionists and the movement of students (within schools and/or districts) caused the need to be flexible in the natural study setting. That flexibility will need to be ongoing. The methodology used in this study, quasi-experimental with equivalent comparison group at the beginning of the study, and the positive results in favor of the treatment condition at the conclusion of the study, qualify this study as ESSA Evidence Level 2.